

Andreas H. W. Küpper

(203) 435-8819 · Cambridge, MA · ahwkuepper@gmail.com
ahwkuepper.info · github.com/ahwkuepper · linkedin.com/in/ahwkuepper

Skills

Programming: Python, C, JavaScript, SQL, Fortran, R, Octave/Matlab, Shell scripting

Tools: Pandas, SciPy, Scikit-learn, NumPy, Matplotlib, Seaborn, Flask, Emcee, OpenMP, MPI

Machine/statistical learning: Logistic/linear regression, Random Forest, SVM, k-nearest neighbors, Bayesian inference modeling, Markov-Chain Monte Carlo, maximum likelihood estimation, bootstrapping/jackknife resampling, KS testing, minimum spanning tree algorithms, Gaussian mixture models, principal component analysis, kernel density estimation

Leadership: Mentored 8 PhD/MSc students, guiding them to publications and conference participations

Communication: 50+ presentations at conferences/public events and 23 [peer-reviewed publications](#)

Experience

Insight Data Science, Boston, MA, *Fellow*

Since 06/2016

- Built [STDand.Me](#), a web application for STD risk assessment using Flask, Bootstrap & D3
- Acquired, cleaned and analyzed CDC and Census data for 3143 US counties with Pandas
- Set up a PostgreSQL database for Census data on ZIP-code level with > 30,000 entries
- Trained and validated various regression models with Scikit-learn to predict STD rates

Columbia University, New York, NY, *Hubble Research Fellow*

Since 09/2013

- Measured the mass of the Milky Way by using Bayesian inference modeling with Markov-Chain Monte Carlo, and compared 10^7 tidal stream models to observational data
- Queried data from the SDSS database and analyzed it using kernel density estimation
- Studied dark matter clumping by producing and analyzing a data set of 10^9 stream stars
- Used unsupervised learning (PCA, Gaussian mixture models) for stream classification
- Organized several large international meetings and workshops (> 100 participants)

Yale University, New Haven, CT, *Research Fellow*

2013

- Developed a Bayesian framework in Python/C for statistical modeling of tidal streams

Universität Bonn, Germany, *Postdoctoral Researcher*

2011 – 2013

- Created a now widely used [algorithm](#) for efficient modeling of tidal streams, which reduces the time for model generation from days to seconds
- Performed maximum likelihood estimation on noisy telescope data
- Detected clumping in star cluster data using minimum spanning tree algorithms

Universität Bonn, Germany, *Graduate Student Researcher*

2007 – 2011

- Studied the formation of tidal streams with high-performance N-body simulations
- Wrote a C library for the cleaning and analysis of large N-body datasets (many GB)
- Extracted insights from simulation data using linear/non-linear least-square fitting, k-nearest neighbor algorithms, KS testing and bootstrapping/jackknife resampling
- Published a popular open-source [C/Fortran code](#) for generating star cluster models

Education

Universität Bonn, Germany, PhD in Astrophysics, *summa cum laude*

2011

Penn State University, PA, Bootcamp on Astrostatistics with R

06/2010

Universität Bonn, Germany, Diplom in Physics (MSc equivalent)

2007